

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Hopkinson, Alan (1999) Tradiciniai komunikaciniai formatai : MARC anaipol ne seniena.
Bibliografija '98, 98 . pp. 21-27. ISSN 1392-1991 [Article]

Published version (with publisher's formatting)

This version is available at: <https://eprints.mdx.ac.uk/7240/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

Tradiciniai komunikaciniai formatai: MARC anaipol ne seniena

Alan HOPKINSON

Midlseksio universitetas, GB - Bouds Green Road, London, N11 2NQ, el. paštas: A.Hopkinson@mdx.ac.uk

1. Aptarimo subjektas

Tikiuosi įrodyti, kad dvi šeimos, tradicinių ir šiuolaikiškesnių formatų, yra ne priešingos, o viena kitą papildančios. Tačiau pirmiausia būtina paaiškinti, kodėl iš viso kilo mintis gretinti tradicinius ir naujus formatus.

Pasistengsiu atsakyti į du aktualius klausimus: Ar mums vis dar reikia MARC formatų? ir Ar MARC varžosi su kitais formatais dėl to paties vaidmens?

Tačiau pirmiausiai reikėtų aptarti tai, ką šiame kontekste mes laikome „formatu“. Formatas yra duomenų perkėlimo automatizuotose sistemose galimybė. Tai gali būti dvi automatizuotos bibliotekinės sistemos (viena iš jų galbūt su paskirstyta duomenų baze) arba tinklo serveris ir asmeninis kompiuteris, turintis prieigą prie serverio, arba asmeninis kompiuteris, kuriuo sudaroma bibliografija, ir nuotolinis kompiuteris serveris, kuriame sukaupti bibliografiniai įrašai.

Formatas yra duomenų struktūrą ir turinį nustatančių taisyklių rinkinys. Kai kurie formatai nustato tik struktūrą, kiti - ir struktūrą, ir turinį. Struktūrai nustatyti MARC remiasi „ISO 2709: bibliografinės informacijos mainų magnetinėje juostoje formatu“¹. Kadangi ši struktūra yra MARC apibrėžimo sudėtinė dalis, MARC be jos nebūtų MARC. Tačiau daugelis kataloguotojų laiko save MARC ekspertais, nieko neišmanydami apie jo struktūrą! Taigi tai jau yra anomalija. Iš tikrųjų yra daug tokio pobūdžio su MARC susijusių anomalijų. Ką tik minėjau „mainus magnetinėje juostoje“, tačiau šiandien beveik niekas nenaudoja magnetinių juostų. Tiesą pasakius, ir paties standarto pavadinimas buvo pakeistas 1996 metais, paskutiniosios jo versijos pavadinimas dabar yra „informacijos mainų formatas“, bet daugelis MARC formatų iki šiol to nepripažino, ir kadangi pagrindinės veikėjos šioje srityje - nacionalinės bibliotekos bei stambiosios kataloginių įrašų tarnybos, tokios kaip OCLC JAV ar BLCMP Didžiojoje Britanijoje - vis dar naudoja magnetines juostas tarpusavyje perduodamos įrašus, darbo su magnetinėmis juostomis metodai tebėra reikšmingi.

Terminas MARC apibūdina visą klasę formatų, tokių kaip MARC, UNIMARC ir bet kurių kitų, turinčių bendrą

įrašų struktūrą, pagrįstą ISO 2709 standartu, o tai reiškia, kad jie visi turi trijų skaitmenų laukų žymas ir polaukius.

JAV Nacionalinė informacijos standartų organizacija šiuo metu rengia naują standartą „Z39.80 Įrašų iškrovimo standartinis formatas“. Jame bus pateiktas apibrėžimas struktūrizuotų bibliografinių įrašų, kuriuos galima eksportuoti iš vienos kompiuterinės sistemos ir importuoti į kitą². Tai bus panašu į mainų formatą, tik be magnetinių juostų apdorojimo reikalavimų.

Mes įvedėme formato apibrėžimą. Naujesni „formatai“ apibendrintai vadinami DTA - „dokumentų tipo apibrėžimais“ (DTD - Document type definitions). Šis terminas vartojamas ir SGML (Standartinė apibendrinta žymėjimų kalba - Standard generalized markup language), ISO 8879³. SGML yra kitos rūšies, o ne tradicinis bibliotekos katalogo formatas. HTML, kodavimo, toliau - tinklalapio, „formatas“ (gerai pažįstamas, nes jūsų World Wide Web naršyklė leidžia jums jį akivaizdžiai pamatyti kaip struktūrizuojantį formatą) yra pagrįstas SGML. Be struktūrizuojančių elementų, apribojamų „mažiau“ (<) ir „daugiau“ (>) ženklais, jame yra kodai, pavyzdžiui, HTML „`„HREF=NAME=“`“ ir sutartiniai žymėjimai, kaip duomenų įvedimas kabutėse, pavyzdžiui, „mailto: someone@univ.ac.uk“.

Tarp tradicinių ir kitų aptariamų formatų yra vienas labai svarbus skirtumas: „tradiciniai“, ISO 2709 įrašų struktūra pagrįsti formatai, vartojami beveik išimtinai vien tik bibliografiniams duomenims, net neįtraukiant į juos viso bibliografinių dokumentų teksto, o kitų formatų taikymas bibliografiniams duomenims tėra viena iš daugybės jų taikymo galimybių. Jie vienodai gali būti naudojami tiek išsisiems tekstams, tiek bibliografinėms nuorodom.

2. Ar mums vis dar reikalingas MARC formatas?

Prieš daugelį metų tarp kompiuterininkų buvo madinga manyti, kad jei kokia nors įranga yra senesnė nei dešimties metų, tai ji jau pasenusi. Tai ypač tiko techninei įrangai. Aštuntojo dešimtmečio viduryje minikompiuteriai pradėjo išstumti didžiąsias ESM, o 1995 m. „Pentium PC“ buvo

galingesnis negu aštuntojo dešimtmečio pradžios stacionari ESM. Dabar asmeniniai kompiuteriai pasensta per porą metų.

MARC buvo sukurtas 1967 m., ir jau 1985 m. atsirado ekspertų, kurie sakė, kad jis paseno ir turi būti pakeistas. 1998 m. jis vis dar su mumis, nepaisant kompiuterių kartų kaitos. Ar galime iš to padaryti kokias nors išvadas?

Akivaizdu, kad mums MARC reikalingas, nes jį naudojame; tačiau galima klausti, ar jis gali būti patogesnis, ar jį galima pagerinti? Ar jis nepaseno? Kad atsakytume į šiuos klausimus, turime užduoti kitą klausimą: kodėl jis yra būtent toks ir kuo galėtume jį pakeisti?

2.1. Įvadas į MARC

MARC kaip įrašo struktūra buvo sukurtas įrašų mainams tarp kompiuterinių sistemų, naudojančių tuo metu paplitusią mainų laikmeną - pusės colio pločio magnetinę juostą. Nepaisant to, jog šiandien daugelis įrašų perduodama kitose laikmenose, kaip lankstūs diskeliai arba per Internetą, daug įrašų vis dar perduodama išlaikant originalo įrašo struktūrą, nors juose ir nesilaikoma originalios įrašo išdėstymo struktūros, kuri būtina magnetinėms juostoms. Labai nedaugelis sistemų savo įrašų vidiniam formatui naudoja MARC struktūrą, tačiau ankstyvuojų laikotarpiu, kai dauguma failų buvo laikoma autonomiškai magnetinėse juostose, o ne interaktyviose diskinėse sistemose, įrašai buvo kaupiami MARC formatu.

Be struktūros, MARC vartoja kodus, pagal kuriuos atpažįstami atskiri katalogo bibliografinio įrašo elementai, faktiškai pagrįsti katalogo kortelės „duomenų elementais“, tokiais kaip pradžioje, antraštė, leidimo duomenys, data ir papildomos nuorodinės kortelės asmenvardžiams, kolektyvams ir antraštei kartu su teminiais ir klasifikaciniais duomenų elementais. Šitaip MARC supranta patyręs MARC kataloguotojas.

Visuotinis šių kodų vartojimas įmanomas tik todėl, kad yra nustatyti griežti, plačiai taikomi bibliografinių duomenų užrašymo standartai - tokie kaip ISBD⁴. Metai po metų augant duomenų mainams paaiškėjo, kad reikalinga dar griežtesnė duomenų kontrolė, ir tokios organizacijos kaip Kongreso biblioteka ir Britų biblioteka ėmė rengti autoritetinių įrašų unifikavimo programas. Kongreso biblioteka taip pat anksti pradėjo rengti protokolo Z39.50 vartojimo projektą⁵ ir sujungtų sistemų projektą⁶, kuris įgalino per naktį automatiškai atnaujinti įrašus pagal patikslintus autoritetinius duomenis, naudojant Z39.50 protokolus.

Nuo 1985 m., kai bibliotekininkai ėmė svarstyti, kas galėtų pakeisti MARC, jau buvo įdėta daug darbo į automatizuotų bibliotekinių sistemų kūrimą, daugiausia Jungtinėse Valstijose. Tuo pat metu panašios sistemos buvo kuriamos Didžiojoje Britanijoje bei kitose šalyse. Tačiau devintojo dešimtmečio pabaigoje dauguma sistemų tikėjū

Didžiojoje Britanijoje nebuvo nieko girdėję apie MARC. Jei apie MARC buvo girdėję, jie teigė, kad jų sistemos bus suderinamos su MARC, nors tai ir reikštų, kad jos galėtų įtraukti MARC duomenis į savo duomenų bazes, bet ir ištraukti juos iš ten. Tarp sistemų tiekėjų, išskyrus tuos, kurie patys buvo MARC įrašų tiekėjai, vyravo įsitikinimas, kad daugelis bibliotekininkų nepageidauja sudėtingo MARC, jei ir norėtų į savo duomenų bazes įtraukti jo įrašų turinį tai be įrašų struktūros subtilybių. Jie taip pat užsimindavo, kad MARC išėjo iš mados ir greitai mirs.

JAV buvo kitaip. Beveik visos sistemos, net skirtos mažoms bibliotekoms, buvo suderinamos su MARC: jos galėjo paimti duomenis iš MARC formatų ir atstatyti MARC įrašus iš savo sistemos duomenų tam, kad galėtų juos perduoti kitoms sistemoms. Daugelis jų leidžia kataloguotojams nebūti susipažinusiais su visomis MARC subtilybėmis, bet iš esmės jos yra grindžiamos MARC. Taigi dauguma sistemų ne verčia vartoti MARC, bet šį vartojimą palengvina.

Didžiojoje Britanijoje sistemos, pagrįstos bendru katalogavimu, tokios kaip BLCMP, kaip ir OCLC Jungtinėse Valstijose, iš savo dalyvių visada reikalavo kokybiško MARC katalogavimo, ir šiuose sluoksniuose MARC būtinumo klausimas niekad nebuvo keliamas.

Žinoma, MARC sudaro įrašo struktūrą ir identifikatorių įrašų rinkinį, tačiau apibrėždamas elementus jis remiasi katalogavimo taisyklėmis. USMARC, UKMARC ir daug kitų nacionalinių formatų apibrėždami elementus naudoja Anglijoje, amerikiečių katalogavimo taisyklėmis (AACR)⁷, o tie, kurie nesiremia AACR, naudoja kitas katalogavimo taisykles, kuriose elementų apibrėžimai pagrįsti tarptautiniu standartiniu bibliografiniu aprašu (ISBD). Be AACR anglosaksiškame pasaulyje MARC būtų visai kitoks. Panašiu būdu MARC buvimas užtikrina, kad AACR ir kiti katalogavimo formatai galėtų būti naudojami automatizuotose bibliotekų sistemose. Taigi AACR turėjo labai didelę įtaką MARC, o MARC savo ruožtu labai smarkiai veikė sistemų, pagrįstų AACR, plėtotę, užtikrindamas, kad jų duomenų bazės būtų suderinamos tarpusavyje. Kita lygties pusė yra įrašo struktūra: apibrėžta įrašo struktūra reiškia, kad įrašai, kad ir kaip jie būtų laikomi sistemoje, gali būti išvesti tokios konkrečios struktūros, kurią galėtų „perskaityti“ kitos sistemos. Gana įdomu, kad MARC iš pradžių bibliografiniame įrašė neapėmė visko, kas buvo numatyta valstybinės bibliografijos reikalams: daugelyje nacionalinių formatų jo silpnoji vieta tebėra ne kūrinių, o atskirų vienetų, t.y. egzempliorių duomenys - tai paaiškėja, kai duomenys perduodami iš vienos sistemos į kitą.

Taigi įdomu pažymėti, kad nuo tų laikų, kai buvo imta pranašauti MARC mirtį, jo vartojimas išaugo.

Taip pat įdomu, kad Kalifornijos valstijos universitetas neseniai ėmėsi projekto sukurti vieningą informacijos prieigos sistemą tarp visų asocijuotų universitetų ir jų kompiuterių. „Po išsamių aptarimų projekto vykdytojai nusprendė, kad sistema, pagrįsta atvirais TCP/IP protokolų

standartais telekomunikacijoms Interneto terpėje, HTTP/HTML - Web tinklo kliento su Web serveriais komunikacijoms, USMARC bibliografiniams įrašams ir Z39.50 protokolu informacijos paieškai iš vieno kompiuterio kitame, yra optimali šiuolaikinė visuotinės tiesioginės bibliografinės prieigos technologija. Per pirminį informacijos rinkimo etapą vykdytojai gavo labai teigiamų atsiliepimų už tokį požiūrį iš galimų tiekėjų⁸.

2.2. MARC privalumai ir trūkumai

2.2.1. Privalumai

- MARC vartoja ISO 2709 struktūrą, kuri yra griežtai apibrėžta įrašo struktūra su galimybe pasirinkti, kaip įvesti polaukius bei indikatorius: MARC pasirenka vieną iš standartinių struktūros variantų; kiti trys nėra vartojami ir minimi dėl išsamumo.
- Sistemų projektuotojai susiduria su fiksuota struktūra ir tiksliai žino, ką reikia įdiegti, nors tai atrodo neįprasta programuotojams, nesusipažinusiems su bibliografinių duomenų apdorojimu.
- Šiuo metu pasaulyje sukauptas milžiniškas kiekis bibliografinių įrašų MARC formatu. Šiuo formatu gali keistis duomenimis daugelis sistemų, nes jose yra atitinkamos programinės priemonės. Įrašus galima lengvai transformuoti iš vienos sistemos į kitą, nes jie visi vartoja ISO 2709 formatą. Kai automatizuotos sistemos tiekėjas patobulina savo programinę įrangą, įrašai gali būti eksportuojami iš senosios ir importuojami į naująją sistemą. Šiomis priemonėmis taip pat galima pereiti nuo vienos tiekėjo siūlomos techninės įrangos prie kitos.
- Formatas nėra kieno nors nuosavybė, nors, kadangi jis buvo sukurtas Kongreso bibliotekoje vykdant MARC projektą, todėl turi savybių, įtrauktų dėl konkrečios sistemos projektavimo ir naudotos techninės įrangos ypatybių.
- Formatas atsižvelgia į kintamo ilgio ir pasikartojančių laukų poreikį, kurio daugelis nelanksčių (ir todėl lengvų programuoti) įrašų struktūrų nepatenkina.

2.2.2. Trūkumai

Galima pastebėti ir kelis trūkumus, tačiau jie nesukelia didelių problemų.

- Palyginti su kitomis struktūromis, neturinčiomis sudėtingos žinyno sistemos, įrašo struktūrą sunku „perskaityti“. Šią struktūrą kataloguotojui sukurti tiesiogiai beveik neįmanoma, reikalinga kompiuterio programa. Tai nėra problema, nes duomenų struktūra skirta tik

mainams. Vartotojai mato įrašus savo sistemoje; įrašo struktūra vartojama tik perduodant duomenis tarp sistemų.

- Įrašo struktūrą sunku programuoti; iš tiesų tai nėra sudėtinga, jei programuotojas turi ISO 2709 aprašymą, kuris įtrauktas į išspausdintą MARC dokumentaciją.
- Įrašo struktūroje numatyti valdymo ženklai, kuriuos sunku naudoti (pvz., ASCII31 polaukio identifikatoriui); vietinės sistemos viduje galima naudoti kitokius ženklus, tokius kaip \$, @ ar ^ vietoje ASCII31, ar netgi palengvinti duomenų įvedimą (pvz., atskirus įvedimo ekranus), akivaizdžiai nurodant, kur tų ženklų reikia, jei tai palengvina darbą.
- Su dideliais failais sunku dirbti.
- MARC nepalengvina ryšio tarp atskirų bibliografinių įrašų. Įdomu, kad būtent šia galimybe buvo papildytas antrasis standarto leidimas, ISO 2709-1981, bet jos naudojimas nepaplitęs, nes prieštarauja priimtai katalogavimo praktikai. Ją apėmė UNESCO Bendrasis komunikacinis formatas⁹, panašus į MARC, tačiau daugelis šio formato vartotojų iki šiol jos neįdiegė. Viena iš priežasčių yra ta, kad įrašai siejami su kūrinių, o ryšiai nustatomi per autoritetinius failus bei iš jų sudarytas rodykles, pagrįstas tam tikrais MARC įrašo duomenų elementais, standartizuotais intelektualinėmis pastangomis. Įrašai lieka paskiri, o intelektualinės sąsajos tarp jų sudaromos imant iš jų duomenų elementus ir įvedant juos rodyklėse. Tai iliustruoja UNIMARC autoritetinių įrašų formatas. Kadangi ryšiai tarp įrašų galiausiai atitinka ryšius tarp objektų, tam buvo surasti kiti būdai, tokie kaip lauko 856 vartojimas, leidžiantis prieigą prie kitų resursų, tokių kaip visas tekstas (pavyzdžiui, prie paties objekto, užfiksuoto katalogo įrašu, ar prie objekto kopijos kitu pavidalu, ar prie jo atitikmens).
- Ar MARC iš tiesų yra standartas? Pasaulyje yra daug skirtingų MARC formatų. Tai bene svarbiausia MARC problema, ir todėl jai aš paskiriu išsiaiškinsi.

2.2.3. Ar MARC yra standartas?

Teisinga yra kritika, kad daugybė MARC formatų mažina MARC, kaip standarto, vertę. Jo struktūra iš pradžių buvo priimta kaip Jungtinių Amerikos Valstijų standartas (ANSI Z39.2, dabar NISO Z39.2), po to - kaip ISO. Kai buvo tvirtinamas JAV standartas, nutarta, kad duomenų elementų apibrėžimai bus pateikti kaip priedas, o ne kaip sudėtinė dalis, kadangi pripažinta, jog skirtingi informacinės bendruomenės sluoksniai kels skirtingus reikalavimus duomenų elementų detalėms. Žinoma, tam tikroje srityje tai, ar formatas yra formaliai pripažintas standartas, nebūtinai gali turėti reikšmės. Tačiau įdomu, jog dėl struktūros buvo susitarta ir ji yra naudojama visuotinai, tuo tarpu dėl

duomenų elementų apibrėžimo tokio sutarimo nebuvo pasiekta ir dėl to kilo problemų, aptariamų toliau.

Pradžioje, septintojo dešimtmečio pabaigoje sukūrus MARC, kai kurios nacionalinės bibliotekos nusprendė turinčios padaryti smulkių originalaus Kongreso bibliotekos MARC pakeitimų.

Daugelis skirtumų yra menki ir nekelia daug sunkumų, norint pervesti duomenis iš sistemos, vartojančios vieną formatą, į vartojančią kitą. Tai tokios problemos kaip: ar UKMARC laukas „513 - Reziümė pastaba“ atitinka USMARC lauką „520 - Reziümė ir t.t.“. Jei sutinkama, kad laukai tapatūs, importuojant duomenis iš vienos sistemos į kitą tereikia konvertuoti lauko žymą. Kiti skirtumai yra sudėtingesni, pavyzdžiui, UKMARC asmenvardis rašomas \$aShakespeare\$bWilliam, UNIMARC - \$aShakespeare \$bWilliam arba \$aShakespeare.\$bWilliam, o USMARC - \$aShakespeare, William (pastebime, kad dolerio ženklas iš tikrųjų vaizduoja specialų ženklą ASCII31). Tai reiškia, kad dviženklis kodas viename MARC formate turintis vienintelę reikšmę, kitame formate yra ekvivalentiškas anaipol nevienareikšmei dviejų simbolių eilutei: „kablelis, tarpas“. Antrosios rūšies pavyzdžių yra ir daugiau. Buvo stengiamasi pasiekti harmonijos tarp JAV, Kanados ir britų formatų. Didžiojoje Britanijoje visi tikėjo, kad harmonizavimas reiškia suvienodinimą. Muzikoje, iš kur yra kilęs, harmonija reiškia „vieno sąskambio ir tolesnių sąskambių garsų derinys“ (Oksfordo enciklopedinis žodynas). Taigi harmonizacija reiškia suderinimą, o ne suvienodinimą. Čia kalbame apie duomenis, kuriems daugelis vartotojų kelia didelius vienareikšmiškumo reikalavimus. Kai kuriais atvejais tai pateisinama, tiesą sakant, vien estetiniu požiūriu, bet kitais atvejais, kaip anksčiau minėta kablelio ir kodo alternatyva, duomenys turi būti labai tiksliai apibūdinami, nes jie bus naudojami rodyklėse. Šiuo atveju reikia ne susitarimo, o vieningumo ir vienodumo. Katalogą ištiks katastrofa, jei jo rodyklėje dėl pagrindo slypinčių skirtumų bus visokių variantų Šekspyro, Viljamų:

Shakespeare\$bWilliam

Shakespeare, William

Shakespeare.\$bWilliam

Automatizuotos sistemos šiuos sunkumus gali apeiti, tačiau paprastai jos kuriamos nenumatant, jog teks spręsti problemas, kylančias dėl skirtingų duomenų standartų naudojimo.

R. W. Hillas savo darbe „Patikimo aprašo sudarymas“¹⁰ nurodo: „MARC harmonizavimas iškėlė daug sudėtingų klausimų, susijusių su atskirų vartotojų grupių poreikiais“. Buvo pripažinta, kad USMARC ir Kanados MARC, būdami visiškai harmonizuoti tarpusavyje, yra tik iš dalies harmonizuoti su UKMARC.

Daugelio problemų šaltiniai yra panašūs į minėtus anksčiau. Papildomas polaukio identifikatorius į UKMARC buvo įvestas labai anksti tam, kad suteiktų didesnio lankstumo apdorojant duomenis. Būtų labai brangu, be to

prarastume lankstumą, jei konvertuotume visus UKMARC įrašus ir perdirbtume visus programų paketus. Iš tikrųjų skirtumai yra labai maži, palyginti su duomenų pavidalo skirtumais, kuriuos atrastume, jei būtų galima automatiškai konvertuoti ir lyginti tarpusavyje nuorodas iš įvairių šaltinių arba ne pagal AACR taisykles sudarytus katalogus. Vis dėlto skirtumai darosi juokingi, kai pagalvoji, kad ir USMARC, ir UKMARC duomenys yra suderinami su AACR taisyklėmis, todėl turėtų būti visiškai suderinami tarpusavyje.

Kai informacija keičiamasi tik tarp panašių organizacijų, perduodant duomenis neturėtų kilti problemų, o vienos šalies viduje, kur visos bibliotekos priėmė nacionalinį formatą, problemų nebūna. Jos atsiranda tarptautiniame lygmenyje. Siekiant šias problemas išspręsti, kaip tiltas tarp nacionalinių formatų buvo sukurtas UNIMARC. Buvo nuspręsta, kad pagal Europos Sąjungos remiamą „Telematikos bibliotekoms“ programą vykdomiems tarptautiniams projektams šie skirtumai yra pakankamai svarbūs, kad galėtų kliudyti projektų efektyvumui, todėl iš visų dalyvių reikalaujama vartoti UNIMARC.

Kituose šio rinkinio straipsniuose kalbama apie UNIMARC, todėl aš paminėsiu tik tiek, kad pagrindinė idėja buvo ši: jei pasikeitimas duomenimis vyks UNIMARC formata, tai kiekvienai šaliai reikės tik vieno programų rinkinio - konvertuoti duomenis į nacionalinį formatą ir konvertuoti iš jo.

Tai nepadėtų vartotojui, savo namuose turinčiam Interneto ryšį, kuris ras nepažįstamo formato įrašą per WWW, o kompiuteryje nebus tą formatą atpažįstančios programos. Šiuo atveju duomenų perdavimui geriau tinka naujesni formatai. Tačiau tokiam vartotojui nelengva ir atlikti paiešką keliose sistemose. Tokio pobūdžio paieškas palengvina standartizuotos sistemos, pavyzdžiui, vartojančios protokolo standartą Z39.50. Pastarasis remiasi požymių rinkiniais, priklausančiais nuo indeksų. Skirtingos katalogavimo sistemos leidžia arba verčia pasirinkti indeksams skirtingus MARC laukus. Paieškai keliose sistemose, kurių indeksavimo metodai yra skirtingi, reikalingas standartas, o viena iš protokolo Z39.50 funkcijų būtent ir yra toks standartizavimas. Suradus ieškomus bibliografinius įrašus, reikia juos peržiūrėti. Z39.50 reikalauja, kad sistema atsiųstų įrašą MARC formatu. Norint, kad vartotojo programinė įranga galėtų atpažinti įrašą, reikia žinoti, koks MARC formatas yra naudojamas. Vis dėlto, kadangi įrašo struktūra yra standartizuota, programa, sugebanti perskaityti vieną MARC formatą, gali perskaityti ir kitą ir turėtų sugebėti atvaizduoti visus duomenis, nors ir negalėdama teisingai jų atpažinti ar išdėstyti skyrybos. Dėl to atliekama daug tyrimo darbų, gramatiškai nagrinėjant skirtingų MARC formatų įrašus tam, kad juos būtų galima naudoti sistemose su Z39.50 protokolu. Verta pažymėti, kad net jei MARC vartojimas būtų nuoseklesnis, bet sistemos sudarinėtų rodykles iš skirtingų MARC laukų, vis viena protokolui Z39.50 prireiktų koduoti

požymių rinkinių. Tai rodo, kad duomenys (nors ir koks aukštas būtų jų standartizavimo lygis) nėra vienintelė sistemos dalis, į kurią reikia atsižvelgti, ir kad indeksus, kurie yra mažiau standartizuoti, taip pat reikia turėti galvoje.

Čia mes ir atradome didžiausią MARC silpnę, jo trūkumą, pasireiškiantį tuo, kad egzistuoja skirtingos nacionalinės versijos. Atsižvelgiant į AACR universalumą (šiandien, jei ne tuo metu, kai buvo kuriamas MARC) šis trūkumas yra nepateisinamas ir turi būti pašalintas. Vis dėlto tai nebūtinai yra MARC silpnė. Filosofai samprotauja apie įmanomus pasaulius, ir galima įsivaizduoti tokį pasaulį, kuriame pasisekė suvienodinti MARC visoje pasaulinėje bibliotekų bendrijoje.

Tai leidžia man tikėti, kad nors MARC formatas turi trūkumų ir, kaip daugelis standartų, vartojamas ne taip, kaip derėtų, padėtis būtų daug blogesnė, jei mes iš viso neturėtume jokio MARC standarto. Vilioja mintis, jog ji galima būtų pakeisti kitu formato standartu. Bet kadangi įrašo struktūra čia niekuo dėta, o iš tikrųjų yra viena jo stiprybių, toks pakeitimas padėties nepagerins. Jei mes negalime pasiekti susitarimo dėl duomenų elementų apibrėžimo ir pateikimo, ar yra pagrindo manyti, kad galėtume susitarti dėl bet kurio kito standarto arba kad, tą pasiekę, „harmoningai“ vartotume jį ateityje?

2.2.4. Duomenų mainai XXI amžiuje

MARC buvo sukurtas duomenų standartizavimui, nes standartizavimo reikia duomenų mainams, tad MARC yra mainų formatas. Bibliotekoms bibliografinio įrašo standarto reikia ir kitam tikslui: tam, kad pasiektų pakankamą vidinio nuoseklumo lygį užtikrinant, jog įrašai suprantami skirtingiems vartotojams. MARC tai palengvina, bet iki galo pasiekiamas šio tikslo taikant katalogavimo taisyklių naudojimą. Mainai anksčiau reiškė bendrą katalogavimą, bet šiandien duomenys perkeliami ir tuomet, kai perkeliamos sistemos. Perkėlimas nėra pagrindinis mainų formato uždavinys, bet tokių formatų buvimas perkėlimą palengvina, ir tai yra geras šio formato pašalinis rezultatas. Kai sistemos perkeliamos, būtent katalogų duomenys, nors ir labai sudėtingi, palyginti, tarkim, su išduoties failais, paprastai yra konvertuojami sėkmingiausiai ir efektyviai.

Vis dėlto vartotojai dabar nori kitokiu būdu prieiti prie duomenų iš skirtingų sistemų. Kalbama apie Interneto ir WWW standartus. Ar tai pagrindas daugiau nebenaudoti MARC? Tarp techninio išsilavinimo žmonių, nelabai suprantančių katalogavimo meną, yra paplitusi nuomonė, kad MARC struktūra nėra tinkama WWW tinklui. Teikiami įvairūs pasiūlymai, bet nė vienas jų iš tiesų nėra skirtas parengti baziniam standartui, kuris būtų pakankamai artimas duomenų elementų žinynei ir kaip MARC pasitarnautų standartizuojant milijonus bibliografinių įrašų visame pasaulyje. Jei standartą kurtume dabar, greičiausiai

neturėtume triženklų lauko žymų ir polaukių identifikatorių. Mes įvestume ką nors sudėtingesnio, bet turėtume nepamiršti, kad duomenis reikia įvesti tokiu būdu, kuris tiktų šio pasaulio kataloguotojams, o tai, ką dabar turime, atsirado evoliucijos būdu ir yra gana vienoda struktūra pagal katalogo kortelės, papildytos sąsajomis su bibliografiniais ir autoritetiniais įrašais, pavyzdį.

Vis dėlto ši struktūra bei duomenų elemento apibrėžimas praverčia, mes galime pateikti duomenis ir tais kitais formatais, kurie nėra mainų tikraja prasme.

Anksčiau aš klausiau, ar MARC galima pagerinti: skirtingi nacionaliniai formatai vystosi visą laiką, bet esminiai pokyčiai, tokie kaip įrašų sąryšio loginiai metodai, visada evoliucionuoja iš lėto. Pirmenybė teikiama stabilumui, o ne kaitai, bet, mano nuomone, tai gal ir nėra blogai.

3. Ar MARC ir kiti formatai varžosi dėl to paties vaidmens?

Grįžkime prie klausimo iškelto pradžioje: ar MARC ir kiti formatai varžosi dėl to paties vaidmens?

Iki šiol neteko pavartoti žodžio „metaduomenys“, nors buvo kalbama vien tik apie metaduomenis. Norint suprasti, ką tai reiškia, geriausia pradėti nuo Davido Sokerio „Metaduomenų pradžiamokslio“¹¹, kur rašoma: „Remiantis „ADAM trumpu metaduomenų vadovu“, <URL <http://adam.ac.uk/adam/metadata.html>>, „bendriausias šio termino apibrėžimas [...] yra duomenys apie duomenis - informacija, aprašanti kitą informaciją“. Taigi bet kokio pavidalo katalogas - turinio sąrašas, aprašas, apžvalga, referatas ar rodyklė - yra metaduomenų rinkinys, jeigu tik juo galima naudotis aprašant kitus šaltinius arba dokumentus. Jis galėjo įtraukti ir MARC įrašus. Atkreipkite dėmesį, kad „antraštinis puslapis“ neminimas. Dublin Core naudojamas kompiuterinei medžiagai identifikuoti taip pat, kaip antraštinis lapas vartojamas spausdintai medžiagai. Kadangi jis aprašo duomenis, esančius Internete, ir pats yra prieinamas rodyklių sudarinėtojams Internetu, duomenys tuo pačiu pavidalu (nors ne tuo pačiu loginiu ar net, kaip aš įtariu, fiziniu adresu) gali sudaryti tų duomenų rodyklę. Tai vyksta taip, lyg jūs kopijuotumėt senus išsamaus katalogavimo leidinyje įrašus iš antraštinio lapo antros pusės ir klijuotumėt juos ant katalogo kortelės. Duomenys Dublin Core formatu Web tinklalapyje negali atlikti metaduomenų vaidmens, bet jų tikslas yra padėti lengviau generuoti metaduomenis.

Michael Day iš UKOLN atliko tyrimą¹² Europos Sąjungos Telematikos taikymų programos projektui „BIBLINK: Dublin Core struktūrizavimas į UNIMARC“. Šio darbo pagrindinė išvada mums (kadangi mes lyginame Dublin Core su UNIMARC) yra tokia: UNIMARC apibrėžtas taip, jog Dublin Core duomenys negali būti

konvertuoti į UNIMARC. Pagrindinės problemos susijusios su tuo, kad Dublin Core neskiria individualaus ir kolektyvinio autoriaus, o taip pat ir su koduotų duomenų kūrimu. M. Day daro išvadą, jog reikėtų tam tikro rankų darbo, bet tai brangiai kainuotų. Mano požiūriu, jei norime sudaryti katalogo įrašus, aprašančius kompiuterinius resursus, vien iš HTML metaduomenų, tai stengiamės atlikti tai geriau už tradicinių leidinių kataloguotojus. HTML metaduomenys yra panašūs į „aprašą“, kurį ISBD traktuoja kaip duomenis, perrašytus iš kataloguojamo objekto, tokio pavidalo ir turinio, kaip jame pateikta. O kataloguotojai visada nori dalykiškai sistematinti savo rinkinius pagal vietos reikalavimus. Nelabai tikėtina, kad dalykiniai terminai, kuriais elektroninius dokumentus papildoma jų autoriai, visiškai atitiks jų katalogavimą. Žinoma, Dublin Core galima išplėsti, ir tai jau galėjo būti padaryta, siekiant didesnio suderinamumo su UNIMARC. Bet UNIMARC ir kiti MARC formatai yra priemonė saugoti dviejų rūšių metaduomenis: tų, kurie buvo sukurti iš originalo, ir tų, kurie buvo sukurti intelektualiomis pastangomis, t.y. tai, ką bibliotekininkai vadina autoritetiniais duomenimis. Nors bibliotekininkai, dalykinantys medžiagą, prie kurios jie nori suteikti prieigą, be abejonės, atsižvelgs į dalykinius terminus, kuriuos ras Dublin Core duomenyse, jie nenorės besąlygiškai jų naudoti, nebent juos būtų sukūrusi centralizuota katalogavimo tarnyba, pavyzdžiui, nacionalinė biblioteka.

M. Day tyrimu buvo siekiama parodyti, iki kokios apimties Dublin Core duomenis galima konvertuoti į UNIMARC tam, kad tarnybos, norinčios kataloguoti savo visų rūšių išteklius, galėtų pasinaudoti Dublin Core „identifikatoriais“, kurdamos katalogo įrašus, kuriuos galima būtų įtraukti į bendresnio pobūdžio medžiagos katalogus. Atvirkščias veiksmas tikriausiai nebūtų praktiškai naudingas, nors aš, rengdamasis šiam pranešimui, parašiau tokią programą- šio straipsnio dalį: nėra prasmės UNIMARC (meta)duomenis konvertuoti į HTML metaduomenis. Niekas nesiruošia projektuoti knygos antraštinio puslapio iš katalogo kortelės. Niekam nebus patogu atrasti įrašą kataloge, konvertuoti jį į HTML ir pateikti jį dokumente kaip pradžienį Dublin Core formatu.

4. Išvada

MARC atvedė prie tam tikro sustabarėjimo apdorojant bibliografinius duomenis, o tai savo ruožtu sukūrė stabilią

aplinką sistemų vystymuisi; tai buvo įmanoma tik todėl, kad visą laiką buvo naudojamos AACR ir ši sritis nesikeitė. Tačiau, jei nebūtume turėję tokio visuotinai priimto standarto (nors ir su nedideliais nukrypimais), nebūtume turėję stabilios dabar esančių vietinių, nacionalinių ir tarptautinių tinklų, sistemų ir įrangos terpės, kurioje galime keistis duomenimis ir kurią galbūt laikome savaime suprantama. Kiekvieno standarto prigimtyje slypi paradoksas: standartai slopina naujoves, tačiau tuo pat metu jie įgalina visus vartotojus pasiekti tam tikrą vienodą pagrindą arba lygmenį, iš kurio jie visi kartu gali peršokti į aukštesnį lygį. Mes dar neperšokome, bet visi jau esame vienodame lygyje!

Manau, kad šiuo straipsniu pavyko įrodyti, kad:

1. MARC, nepaisant jo trūkumų, vis dar tarnauja reikšmingam tikslui - katalogų kūrimui. Jo trūkumai iš dalies atsirado dėl to, kad standartas beveik nepakeistas buvo vartojamas ilgą laiką, o tuo metu pasikeitė duomenų perdavimo laikmenos ir būdai. Standartą pakoreguoti buvo galima, bet tai turbūt būtų sukėlę dar didesnius jo vartojimo praktikos skirtumus. Taip pat tai būtų sukėlę problemų, kurių šiuo metu išvengėme, nes turime milžiniškus bibliografinių įrašų masyvus. Kitu atveju būtų sukurtos atskiros skirtingo sudėtingumo lygio sistemos. Automatizuotų bibliotekinių sistemų tiekėjai būtų turėję atsižvelgti į aibę skirtingų formatų. Dabar jie gali projektuoti savo sistemų interfeisus, remdamiesi ISO 2709, ir šiek tiek modifikuoti savo sistemas, atsižvelgdami į nacionalinių MARC formatų skirtumus.

Nepaisant esamų problemų, kurios kai kam gali atrodyti rimtos, daugumai kuklių bibliotekinių sistemų specialistų, triūsiančių, kad parengtą sistemą savo bibliotekų skaitytojams ir prieigą prie išorinių resursų savo kataloguojantiems kolegoms, tereikia žinoti labai nedaug apie šias problemas. Nepaisant to, kas buvo išdėstyta, MARC struktūros griežtumas tarptautiniu mastu, o duomenų elementų tikslumas - nacionaliniu mastu užtikrina, kad duomenų perdavimas iš vienos sistemos į kitą ir paskirstytas katalogavimas yra kur kas sėkmingesnis negu tuo atveju, jei nebūtų tokio standarto kaip MARC.

2. MARC tikslas skiriasi, bet papildoma vėlesnės kartos formatų, tokių kaip Dublin Core, tikslus.

Galbūt MARC ir sensta, tačiau jis vis dar gyvas!

¹ International Organization for Standardization. *Format for Bibliographic Information Interchange on Magnetic Tape*. Geneva, ISO, 1981 (ISO 2709-1981). 3rd edition: *Format for Information Interchange*. Geneva, ISO, 1996 (ISO 2709-1996).

² NISO. *Standard Format for Downloading Records (NISO Standards Committee AJ)*. [WWW] <http://www.niso.org/commitaj.html> (27.04.98).

³ International Organization for Standardization. *Standard generalized markup language*. Geneva, ISO, 1986

⁴ ISBD. *The International Standard Bibliographic Descriptions are cataloguing codes dealing with the description of documents. The main one on which the others are modelled is ISBD(G): General International Standard Bibliographic Description*. Rev. ed. Munich, K.G. Saur, 1992

⁵ *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*. Washington DC, Z39.50 Maintenance Agency, July 1995.

⁶ Dempsey, Lorcan. *Libraries, Networks and OSI*. Westport, Meckler, 1992

⁷ *Anglo-American Cataloguing Rules*. 2nd ed., 1988 rev. London, Library Association, 1988. Looseleaf, updated.

⁸ Pollard, Marvin. *California State University Unified Information Access System: project report*. [WWW] <http://uias.calstate.edu/uiasproject.html>, 9 July 1997 (27/04/98)

⁹ The latest version of CCF is defined in: *CCF/B: the Common*

Communication Format for Bibliographic Information. Paris, UNESCO, 1992

¹⁰ Hill, R.W. *Setting the Record Straight: a Guide to the MARC Format*. 3rd ed. Boston Spa, British Library National Bibliographic Service, 1998

¹¹ Stoker, David. *A Beginner's Guide to Metadata*. *CIQM Newsletter* 3(1) January 1998

¹² Day, Michael. *Mapping Dublin Core to UNIMARC*. [WWW] http://www.ukoln.ac.uk/metadata/interoperability/dc_unimarc.html (25/04/98)